DOCUMENT RESUME

ED 308 691 FL 017 989

AUTHOR Cohen, Andrew D.
TITLE The Role of Instructions in Testing Summarizing Ability.
PUB DATE 89
NOTE 21p.; Document contains small print.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Answer Keys; Comparative Analysis; *English (Second Language); Foreign Countries; *Hebrew; Higher Education; Item Analysis; Language Tests; Second Language Instruction; Testing Problems; *Test Theory; Uncommonly Taught Languages; *Writing Processes
IDENTIFIERS *Instructions; Israel; *Summarization

ABSTRACT

A study investigated the effects of specific guidelines in the taking and rating of tests of summarizing ability. The subjects were 63 native-Hebrew-speaking students enrolled in English-as-a-Second-Language (ESL) courses at the Seminar Hakibbutzim Teacher Training College in Tel Aviv (Israel). The subjects were given two sets of instructions (guided and minimal) for summarizing texts in both English and Hebrew. The guided instructions gave specific directions on how to read the texts and on how to write the summaries, while the minimal instructions gave traditional minimum direction. Two scoring keys were developed from the summaries of nine Hebrew-speaking and nine English-speaking experts, respectively. The summary guidelines developed in this case did not make a sizable contribution to summary development, but an item-by-item analysis revealed that the guided instructions were both helpful and detrimental. In some cases, the guided instructions helped respondents to identify key elements to summarize, but in other cases, they may have dissuaded the respondents from including details that in fact proved essential in the eyes of the experts upon whom the rating key was based. (Author/DJD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED308691

The Role of Instructions in Testing Summarizing Ability

Andrew D. Cohen, School of Education, Hebrew University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Cohen, A

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FL017989

BEST COPY AVAILABLE

THE ROLE OF INSTRUCTIONS IN TESTING SUMMARIZING ABILITY[1]

Andrew D. Cohen, School of Education, Hebrew University

ABSTRACT

The main purpose of the study was to determine the effects of specific guidelines in the taking and the rating of tests of summarizing ability -- tests in which respondents read source texts and provide written summaries as a measure of their reading comprehension level as well as of their writing ability. Other issues under consideration in this study included similarities and differences in native vs. foreign-language summaries, the foreign-language proficiency level of the summarizer, and the relationship between the language in which the summary was written and the native language of the rater.

The subjects for this study were 63 native-Hebrew-speaking students from the Seminar Hakibbutzim Teacher Training College in Tel Aviv. Twenty-six were from two high-proficiency English-foreign-language (EFL) classes and 37 were from two intermediate EFL classes. Four raters assessed the students' summaries in the study. The two who rated the Hebrew summaries of the Hebrew texts were both native Hebrew speakers. Of the two rating the Hebrew summaries of the EFL texts, one was a native Hebrew speaker and the other an English speaker. Five texts were selected for the study, two in Hebrew and three in English (each intermediate and advanced EFL student being asked to summarize two). Two sets of instructions were developed. One version was "guided" with specific instructions on how to read the texts and how to write the summaries. The other version had the typical "minimal" instructions. The scoring keys for the texts were based on the summaries of nine Hebrew-speaking and nine English-speaking experts respectively. All 63 respondents summarized the first Hebrew text, 53 summarized the second Hebrew text, and on the average, slightly more than a third of the students wrote summaries for the EFL texts.

The study did not demonstrate that the guidelines developed in this case made a sizable contribution. Yet they did seem to have a certain influence. An analysis of summaries on an item-by-item basis revealed that the guided instructions appeared to be both helpful and detrimental. In some cases, they assisted respondents in finding the key elements to summarize, and in other cases they probably dissuaded the respondent from including details that in fact proved to be essential in the eyes of the experts upon whom the rating key was based.

---

[1]I wish to acknowledge Elana Shohamy for her assistance in the design of this study and Rachel Segev for her help in collecting the data. I would also like to acknowledge Don Porter, Gissi Sarig, and Iris Geva for their constructive comments regarding the study. In addition, I wish to thank the eighteen informants who graciously consented to provide expert summaries of the texts.

Summarization tests are complex in nature. In order to perform successfully on such tasks, test takers need a considerable amount of information -- information that often is not provided. Test takers need to know the best way to read the given text(s) in order to prepare the best written summary for the given audience. The preparation of a summary outside the framework of a test is a difficult enough task; summarizing for test purposes compound the difficulty. In summarizing, the reading portion entails identifying topical information, distinguishing superordinate from subordinate material, and identifying redundant and trivial information. The writing up of the summary entails the selection of topical information (or generating it if it is not provided), deleting trivial and redundant information, substituting superordinate material or macro-propositions, and restating the text so that it is coherent and polished (Kintsch & van Dijk 1978, Brown & Day 1983, Chou Hare & Borchardt 1984).

In a study of high- and low-proficiency college-level EFL students, Johns and Mayes (forthcoming) found neither group to be using macro-propositions or meta-statements in their summaries. Furthermore, the low students were found to be doing a considerable amount of direct copying of material from the source text into their summaries, as it was required that the summaries be written in the second language as well. In an earlier study, Johns (1985) had found that underprepared natives were likewise more prone to use reproductions (copying and paraphrase) in their summaries than macro-propositions. In a study of fourteen English-speaking beginning and intermediate undergraduate students of French, Cumming, Rebuffot, and Ledwell (1989) found that both the beginners and intermediates used equivalent propositions of high-order problem solving strategies while reading first- and foreign-language texts and while writing summaries in both languages. The key variable was the level of "literate expertise in mother tongue" -- those more expert tended to use the high-order thinking strategies more (e.g., holistic rather than fragmented understanding of text).

In a case study, Sarig (1988) helped explain the propensity to lift material directly out of the text for use in summaries. She found conceptual transformation or reconceptualization at the macro-level to be a skill that did not come naturally either in native or foreign-language summarizing for a competent college student. She concluded that it had to be taught explicitly. Nonetheless, the explicit teaching of such reconceptualization may not yield such positive results either. Bensoussan and Kreindler (in press), for example, found that whereas EFL students with a semester's training in summary writing now saw summaries as an important tool for grasping the gist of a text, they still expressed frustration at their inability to distinguish macro- from micro-propositions.

Hence, summaries are problematic and yet they keep appearing in tests of reading and of writing, largely for their appeal as "authentic" tests, since people need to summarize in the real world. Of course, a real-world summary may be quite different from a test summary. Real summaries are usually prepared for someone who has not read the text and simply wants to know what it is about. Test summaries are prepared for an assessor who has already decided what the text is about and wants to see to what extent the respondents approximate these prior decisions.

Aside from the problem of a summary task as a valid measure of reading comprehension, there is a further problem of reliability

regarding the ratings of the written summaries. It is perhaps not surprising that the statistical results from summarizing tasks are not always consistent with results from other types of tests (e.g., multiple-choice, short-answer, and cloze). Shohamy, for example, set out to compare tests of summarizing English-foreign-language (EFL) texts to tests with a multiple-choice and an open-ended response format -- with responses either in native language or foreign language, depending on the test version. She found the results from the summarizing data so inconsistent with the results on the other subtests that she eliminated the findings from the published study (Shohamy, Personal Communication, and Shohamy 1984).

A previous study by this investigator (Cohen, forthcoming) had as its purpose to investigate the ways in which respondents at different proficiency levels carry out summarizing tasks on a reading comprehension test and how raters deal with the responses. The respondents for that study were five native Portuguese speakers who had all recently completed an EAP course (at the Pontifícia Universidade Católica de São Paulo, Brazil) with an emphasis on reading strategies, including summarizing. They represented three proficiency levels. Two EAP course instructors who typically rated the EAP exams of summarizing skill at the PUC also participated in the study as raters.

The findings showed that the respondents in that study had little difficulty identifying topical information, yet they had difficulty in distinguishing superordinate, non-redundant material from the rest, due in large part to an insufficient grasp of foreign-language vocabulary. For their written summaries, they did not need to generate topic information because all the texts provided it. They did not, however, have a good sense of balance with respect to how much to delete. Either they were too vague and general or too detailed. While there was some concern for coherence production, there appeared to be relatively little attention paid to producing thoroughly coherent and polished summaries. In essence, the respondents in the study appeared to be more concerned about their interpretation of the source text than they are about their production of a summary. Furthermore, the inconsistencies in the raters' behavior underscored the importance of developing rigorous rating keys with main ideas and connecting schemata for each text.

The main purpose of the present study was to determine the effects of specific guidelines in the taking and the rating of tests of summarizing ability -- tests in which respondents read source texts and provide written summaries as a measure of their reading comprehension level as well as of their writing ability. Tests of summarization do not usually include specific, guided instructions for the respondents as to how to construct their summary. Although it may be thought that respondents would draw upon previous knowledge as to how to summarize -- i.e., from classroom instruction, it would appear that they do not necessarily exercise this knowledge when in a testing situation. In addition, the raters may also lack specific enough guidelines in order to rate the summaries reliably. Other issues under consideration in this study included similarities and differences in native vs. foreign-language summaries, the foreign-language proficiency level of the summarizer, and the relationship between the language in which the summary was written and the native language of the rater.

The study asked the following questions:

1. In what ways do guided instructions affect performance on a summary task?
2. To what extent does the language of the source text play a role when the summary is written in the respondent's native language?
3. What is the relationship between indices of EFL proficiency (i.e., EFL course level/course grade/university entrance exam score) and ability to summarize in a foreign and a native language?
4. How does the native language of the rater influence their rating performance on the summaries, and how consistent are the ratings across raters?

Design of the Study

Subjects

The Respondents: The subjects for this study were 63 native-Hebrew-speaking students from the Seminar Hakibbutzim Teacher Training College in Tel Aviv. Twenty-six were from two high-proficiency English-foreign-language (EFL) classes and 37 were from two intermediate EFL classes.

The Raters: Four raters assessed the students' summaries in the study. The two who rated the Hebrew summaries of the Hebrew texts were both native Hebrew-speaking undergraduates in their last year at the Hebrew University. Of the two rating the Hebrew summaries of the EFL texts, one was a native-Hebrew-speaking freshman at the Hebrew University and the other an English speaker doing a graduate degree at Tel-Aviv University. The latter was highly proficient in Hebrew, as she had received twelve years of schooling at a Hebrew day school in the U.S., was married to an Israeli, and had lived in Israel for five years.

Instruments

The Texts: Five texts were selected for the study, two in Hebrew and three in English. The two Hebrew texts were intended to reflect two levels of difficulty -- both in terms of content, complexity of language, and summarizability (i.e., how easy they appeared to summarize). The first Hebrew text (800 words) was entitled, "The Film Media -- from a Play Thing to an Art Form." It was divided into five sections, each containing a subtitle, thus making it "summarizer-friendly." The second text (1,200 words) was entitled, "The Problematicity of Modern Israeli Prose -- from Detachment to Continuity." This text was written in a more problematic style, and a successful summary of the text required a certain degree of background knowledge about the topic.

Three EFL texts were selected so as to represent three levels of difficulty. The easiest (1,000 words), entitled "How to Avoid Foolish Opinions," presented a series of procedures for avoiding such opinions. Its clear organization facilitated summarizing. The second text (1,200 words), entitled "Modern Constitutions," was more complex both in language and in its conceptual organization. The third article (850 words), entitled "Specialization," was intended to be the most difficult article, but actually turned out to have a relatively easy structure for purposes of summary.

Instructions for Summarizing: Two sets of instructions were developed. One version was "guided" with specific instructions on how to read the texts and how to write the summaries. The other version had the typical "minimal" instructions. The guided instructions told the respondents to read each text in order to identify the most important points -- those that contained the key sentences in each

paragraph or those that would make the summary interesting to read. The respondents were then instructed to write the summary such that the content would be reduced to only the essential points and that less important details or those detracting from the main points would be eliminated. They were also requested to write briefly -- e.g., 80-120 words per summary.[2] Their text was to comprise one paragraph with all the ideas linked by connecting words. They were requested to write the summaries in their own words, and in the case of Hebrew summaries of EFL texts, not to translate word for word. They were also asked to write a draft first and then to copy it over legibly (see Appendix A).

The more traditional, non-guided instructions simply told the respondents to read each text so as to be able to write a summary of it. They were asked to be brief -- 80-120-word summaries, to write their summary first in draft form and then to copy it over legibly.

Construction of Rating Keys: The two Hebrew texts were read and summarized by nine Hebrew speakers, all experts in the areas of reading and writing. Three were university lecturers, specializing in discourse. The rest were university students of language arts and teachers in their own right. The three EFL texts were read and summarized in English by nine native English-speakers, most of whom were university instructors of EFL. The summaries of these experts were analyzed and a key was constructed to include only those main ideas and linking ideas that a majority (i.e., five or more) of the experts had included in their summaries (after Sarig 1989). In this study, no effort was made to distinguish the macro- from the micro-propositions in the scoring key.

The key for the Hebrew texts was produced by one of the two raters. The key for the EFL texts was produced by the investigator and then translated into Hebrew by the Hebrew-speaking rater of the EFL texts, whose English skills were also quite advanced. The EFL key was translated into Hebrew. The keys appeared in the form of a list of numbered ideas, each in sentence form. Appendix B shows the level of agreement for each point selected per passage.

The summaries of the Hebrew experts reflected an 80% average agreement as to which main ideas and connecting ideas should be belong in the summary. The summaries of the EFL experts reflected an 85% average agreement as to which ideas to include. Hence, it should be noted that there was not full agreement even among experts as to which ideas were essential to the construction of a meaningful summary.

Criterion Measures: Measures also included two pieces of background information on all the respondents -- the score that they received on the Psychometric (college-entrance) Exam and the mid-year grade that they received in their EFL course.

Data Collection Procedures

The students were requested to write the summaries in two meetings of their EFL course -- the first one for the Hebrew texts, the second one for the English texts. The sessions took place in December, 1988, and were one week apart. Each lasted approximately 1½ hours. The task with guided instructions was administered to every other student according to how they were seated in a given classroom. The other students received the unguided version. All students

---

[2]The stipulation of number of words is based on common practice in the United States. In Israel, students are used to being told to write a paragraph or a page (a less precise measure).

received in the second sitting the same type of instructions as they had received in the first, as this time their name was written on the instruction sheet in advance. The tasks were not presented as obligatory and were not to count towards the students' grade in the course. Fifteen percent of the students did not summarize the second Hebrew text because they found it too difficult, and about 65% did not summarize the EFL texts for a similar reason.

All 63 respondents summarized the first Hebrew text, 32 receiving guided instructions and 31 not. Fifty-three respondents summarized the second Hebrew text, 27 receiving instructions and 26 not. On the average, only a little more than a third of the students wrote summaries for the EFL texts. A breakdown of the numbers of respondents for each text according to version (guided vs. unguided) and respondents' level of proficiency (intermediate vs. advanced) appears in Table 1.

Although the instructions for both the guided and unguided versions specified that the students were to write a draft summary and then copy it over neatly, only seven of the 67 students did this for the Hebrew summaries, four of the intermediate students and three of the advanced ones. As for the summaries of the EFL texts, only one advanced student wrote a draft, as that student had done for the Hebrew texts as well. Thus, this element in the instructions was not followed in either of the two versions.

Data Analysis Procedures

The raters received brief training by the investigator in rating the summaries and then did several of the ratings in the presence of the investigator in order to resolve problems of immediate concern. The raters were also asked to make note of problems to be resolved in consultation with the investigator after completing the ratings. Each main idea and linking idea received one point in the rating process. No effort was made to have the pairs of raters of the Hebrew and EFL texts "conference" together while learning how to do the ratings. However, ratings of the two pairs of raters for the Hebrew and EFL texts were correlated to determine interrater reliability and discrepancies were identified on a point by point basis in order to determine the types of ideas for which raters had difficulty reaching consensus.

Crosstabulations and ANOVA were run using the SPSS-PC program to check for differences in success at summarizing between the respondents given the guided and unguided versions, and between those at the intermediate and advanced levels. The Correlations program was run to determine interrater reliability and to assess the relationship between summarizing and other variables -- i.e., performance on the psychometric exam (an entrance examination for Israeli universities) and the respondents' first-semester grade in their EFL course.

Findings

1. The Effect of Guided Instructions on Summarizing

a. The Summaries of Hebrew Texts: Overall, guided instructions did not seem to enhance summarizing. In fact, on the first Hebrew text, the one on films, the second rater rated the unguided group significantly better than the group with guided instructions. Yet when behavior was examined at the idea-by-idea level, some interesting differences emerged suggesting that the provision of guided instructions has a differential effect on

summarizing, perhaps helping in some instances and interfering in others. In examining the main ideas within that Film text, item #5 was one for which the unguided group were rated significantly better by both raters (see Table 2). This idea involved giving details as to why film is an art form (e.g., it deals with shape, color, movement, words, and music). It is likely that those getting the instructions warning them not to include unnecessary details were reluctant to provide this level of detail for fear that it would be rated as extraneous.

On the second Hebrew text there were two significant differences by version, one favoring the guided and one favoring the unguided instructions. In the first instance, the guided group which was told to identify all the main ideas in the key sentences included idea #6 more than did the unguided group. This idea consisted of the second of two examples of "detachment and then reintegration in Israeli prose." In other words, the guided group was more sensitive to including both of the examples. In the second instance, the unguided group was more likely to include in their summary a linking statement to the effect that modern Israeli prose is characterized by continuity in the midst of apparent separatism (see Table 2). In this case, it was the guided group that was reminded of the importance of linking up ideas and yet the unguided group practiced it more successfully in this case.

b. The Summaries of EFL Texts: In the EFL text summaries, there were two instances of significant difference by version, according to the first rater. The guided group were more likely to include the first idea in the summary of the Modern Constitutions text -- namely, the idea which provides an historical perspective for the passage. Then the unguided group was more likely to include the detail that countries differ as to the number of special checks and balances stipulated by their constitution (see Table 2). Hence, the group given the fuller instructions seemed more sensitive to include the introductory idea while they once again appeared to be reluctant to include detail that might not be deemed central to the summary.

2. The Language of the Text and Native-Language Summaries

In this study, performance in summarizing native-language texts appeared superior to that of performance on EFL text summaries. Just considering those subjects who wrote summaries in both languages, the overall mean for the summaries of the Hebrew texts was significantly higher than the overall mean for the summaries of the EFL texts averaged across raters (see Table 3).[3] Thus, the writing of native-language summaries for the EFL texts could be construed to be a more difficult task than writing summaries for the Hebrew native-language texts. For one thing, writing the summary in the native language eliminated the possibility of lifting material directly from the passage without fully understanding it, thus making the summarizing task more difficult.

While the two raters of the EFL text summaries had similar average total ratings for the three texts separately and overall, the two raters of the Hebrew texts differed, with the first rating more leniently (see Table 3). In addressing question #4, below, we will

---

[3]One analysis still remains to be done, namely, to submit the results by main idea in both languages to a Rasch analysis to determine whether these items fall on the same unidimensional scale.

look at the main ideas that raters characteristically agreed upon and disagreed upon in their ratings.

3. The Relationship between Indices of EFL Proficiency and Summarizing Ability

It was found that respondents' EFL course level was not related to their Hebrew summarizing ability on an overall basis. Both the intermediate and the advanced groups scored similarly on the average, according to both raters (see Table 4). On an item-by-item basis, however, there were several significant differences on the Film text. The second idea in the key described the socioeconomic status of the audience that used to come to the picture shows of old. Both raters judged the intermediate respondents as including this idea significantly more often, and the second rater assessed the advanced respondents as more likely to include the next-to-last idea which involved specifying the ways in which modern movie technology have turned films into an art form (i.e., selection, focus, and editing) (Table 5).

With respect to the EFL summaries, however, the seventeen intermediate-level students who wrote summaries for the EFL texts outperformed the ten advanced students who responded (see Table 4). It must be noted that the advanced students had one essay in common with the intermediate students (the one on Constitutions) and one advanced article (on Specialization) which their mean was based on, while the intermediate students had the one article in common and an easier one (on Foolish Opinions). This can help explain the difference in favor of the intermediate group.

On an item-by-item basis, only the third idea on the Constitution text showed significant difference. It involved a justification for why the constitution was put above other laws. A third of the intermediate respondents (4 out of 12) included this point in their summaries while none of the students in the advanced EFL course (10) did (Table 5). It would stand to reason that the more advanced students would include this point, but then the issue of what makes someone an advanced student is called into question. It may be that advanced students know more grammar and vocabulary, but do not necessarily have better cognitive powers of discernment and choice when it comes to summarizing a complex text.

The respondents' semester grade in their EFL course and their EFL grade on the university entrance exam (the Psychometric) were also correlated with summary performance. The results produced no significant correlations whatsoever. Such findings would suggest that the cognitive abilities called for in summarizing are not necessarily the same ones needed to obtain high grades in the EFL courses or do well on the EFL subtest of the entrance exam.

4. The Effect of Rater's Native Language on Summary Ratings and Rater Interreliability

In answer to the second question above, it was noted that the two raters of the EFL texts had similar total ratings ($r = .85$, $p<.001$), although the first was a native Hebrew speaker and the second a native English speaker. Yet when we look at the level of consistency from idea to idea within the three EFL texts, we note that there are marked differences (Table 6). On the Opinions text, the raters agreed on their ratings of the fifth and sixth ideas, and more or less so on the third idea, but not for the first or opening statement (i.e., the article suggests procedures for avoiding foolish opinions), the

second, or the fourth (an extension of the second suggestion).

On the second text, the raters agreed on the opening statement which gives an historical perspective and on three of the basic points (i.e., a justification for constitutions being above law, protection of special communities, constitutions of countries differ), but disagreed on the topic sentence linking the passage together cohesively (in terms of the constitution as a fresh start) and on an item dealing with details of what rights a constitution includes. On the third text, the raters agreed on all but the opening idea (i.e., that by the end of the nineteenth century the intellectual generalists had given way to specialists).

Thus, the raters were inconsistent in their ratings of several of the more global, linking ideas and of ideas involving details in the EFL texts. These discrepancies would not appear to be the result of native language differences. Instead, it would seem that these are perhaps the ideas that lend themselves to the most controversy in rating, even when a precise key is provided.

For purposes of comparison, let us look at the intercorrelations of the ratings of the two Hebrew native speakers on the two texts that they rated. On the Film text, there was significant agreement on five of the points, four involving basic description and one involving basic exemplification (i.e., what makes film an art form) (Table 6). There was inconsistency on three points, two involving contrast (both involving the contrast between then and now) and one involving detailed exemplification (i.e., regarding modern technology in moving making).

On the Modern Israeli Prose text, there was considerable agreement involving the opening contrastive idea (i.e., separatism and integration), two examples of this, and a linking idea (i.e., dealing with the continuity in Israeli prose) (Table 6). There was some agreement concerning a descriptive point regarding the origins of separatism, and less agreement regarding two other descriptive points -- a point about continuity and a point introducing the two examples of separatism and integration.

This study did not attempt to investigate why raters disagreed on their assessments of respondents' answers as conducted in a previous study (Cohen 1988), but the findings would indicate that certain ideas on a rating key are problematic for scoring and others are not, and that the language of the rater may not be as important here as other factors.

Discussion and Conclusions

This study had as its main focus to investigate the contribution that instructions make to a test of summarizing ability. Previous research on test-taking strategies would suggest that instructions on tests may not even be read if they are heeded to at all (e.g., Cohen 1984). The intention in this study was to build into the test a set of instructions that would serve as a genuine guide for the respondent. It had been the experience of the investigator and of colleagues that test instructions were often notoriously vague and that they presupposed an understanding of how to do the task -- especially in the case of an activity as complex as summarizing.

The current study did not demonstrate that the guidelines developed in this case made a sizable contribution. Yet they did seem to have a certain influence. An analysis of summaries on an item-by-item basis revealed that the guided instructions appeared to be both helpful and detrimental. In some cases, they assisted respondents in

finding the key elements to summarize, and in other cases they probably dissuaded the respondent from including details that in fact proved to be essential in the eyes of the experts upon whom the rating key was based.

In addition, performance in summarizing native-language texts was superior to that of performance on EFL text summaries. It was also found that respondents' EFL course level was not related to their Hebrew summarizing ability on an overall basis. Both the intermediate and the advanced groups scored similarly on the average, according to both raters. Finally, with respect to interrater consistency, it was found that the raters differed in their ratings of several of the more global, linking ideas and of ideas involving details in the EFL texts. These discrepancies would not appear to be the result of native language differences. Instead, it would seem that these are perhaps the ideas that lend themselves to the most controversy in rating, even when a precise key is provided.

Limitations and Suggestions for Further Research

Although the intention was to indicate for each summary who the text was being summarized for, this was inadvertently left out of the instructions for both groups. If we take the EFL text on Specialization, for example, it was intended that the respondents would be told, say, that they work for a company that makes documentary movies and that their boss has asked them to read an article on the dangers of overspecialization and to summarize it for him in order that they may get some ideas for the preparation of the script for the documentary. Perhaps the addition of this modicum of functionality would have prompted more of the students to do the task in a context where participation was not obligatory.

It may also have been beneficial to the research to have to summaries count for course credit. Then perhaps the students would have taken the EFL text summaries, and even the second Hebrew text summary, more seriously. The fact is that the respondents may not have behaved on these tasks as they would on a genuine test.

The first rater of the Hebrew texts had an observation regarding the scoring key. She felt that by using a key based on the judgment of "experts," it was skewing the assessment away from the level of the students who were being assessed. She felt that the experts wrote a different type of summary from the seminar students -- more of a logical sequence in the ideas, a tighter use of words, and a greater quality of writing. Besides for these observations, she noted that there were points that the experts did not agree upon and so they were left out of the key. Yet students felt these points to be important enough to include. A compromise would be to build a rating key based both on the suggestions of the respondents as to key ideas and the insights of the examiners, as was done by Bensoussan and Kreindler (in press).

In the work by Bensoussan and Kreindler, as soon as the respondents finished summarizing the given texts, they were presented with the two researchers' set of main ideas for the summary arrived at through conferencing, and were asked as a group to react to that set. If they disagreed with any points, they had to convince the teacher researchers that a change was in order in the scoring key. Apparently the students became more proficient at this task as they performed it various times. While at the beginning of the course the teachers dictated the correct summaries, by the end of the semester, the students had learned how to suggest changes (Bensoussan, Personal

Communication).

Finally, no test-taking data were collected in this study as had been collected in studies reported on elsewhere (e.g., Cohen 1984, 1988). Hence, there was no assurance, for example, that the students taking the guided version paid full attention to the elaborated instructions that they received. It was impossible to read those instructions aloud since every other student received the unguided instructions. Perhaps a future study would collect data regarding the processing of the instructions. Furthermore, a follow-up study would include rater conferencing with other raters, say in pairs, in order to allow for potentially greater consistency across raters. Such conferencing could also be studied at the process level through analysis of verbal reports in order to determine what such conferencing entails. Future work could also distinguish main ideas at the level of micro- and macro-propositions to see the extent to which the level of abstraction of the proposition influences the behavior of the rater.

Table 1

Respondents for Text Summaries

	Guided Version			Unguided Version			Overall
EFL Level:	Int.	Adv.	Total	Int.	Adv.	Total	
Hebrew Texts:							
Film as Art	19	13	32	18	13	31	63
Modern Israeli Prose	16	11	27	17	9	26	53
EFL Texts:							
Foolish Opinions	12	-	12	5	-	5	17
Modern Constitution	8	3	11	4	7	11	22
Specialization	-	3	3	-	4	7	10

Table 2

Summaries of Hebrew and EFL Texts by Version
(with/without Guided Instructions)

	Rater 2 - Total Perf. Version		Rater 1 - Idea 5 (Details) Version			Rater 2 - Idea 5 (Details) Version		
	Points	Guided	Unguided	Pts.	Guided	Ung.	Pts.	Guided	Ung.
	0	-	-						
Film	2	4	1	0	26	15	0	28	18
as	3	13	2	1	6	16	1	4	13
Art	4	8	18						
Text	5	4	7						
in	6	2	1						
Hebrew	7	1	2						
	N = 32		31						
	d.f. 5			d.f. 1			d.f. 1		
	Chi-Square=15.19**			Chi-Square=7.48**			Chi-Square=6.92**		

	Rater 1 - Idea 6 (Second Example) Version			Rater 2 - Idea 7 (Linking Idea) Version		
	Pts.	Guided	Ung.	Pts.	Guided	Ung.
Modern	0	7	14	0	15	7
Israeli	1	20	12	1	12	19
Prose						
Text	N = 27		26			
in	d.f. 1			d.f. 1		
Hebrew	Chi-Square=4.32*			Chi-Square=4.47*		

** p<.01
* p<.05

	Rater 1 - Idea 1 (Perspective) Version			Rater 1 - Idea 6 (Detail) Version		
	Pts.	Guided	Ung.	Pts.	Guided	Ung.
Modern	0	0	5	0	6	10
Consti-	1	11	6	1	5	1
tution						
Text	N =	11	11			
in	d.f. 1			d.f. 1		
EFL	Chi-Square=6.47**			Chi-Square=3.67*		

** p<.01
* p<.05

Table 3

Ratings by Language of Text and by Rater

	N	Hebrew Texts Ave. % Score	S.D.	EFL Texts Ave. % Score	S.D.	t value	d.f.
Raters & Texts Combined	27	51.73	14.06	35.08	14.86	4.78***	26
Overall Hebrew Texts							
Rater 1	63	56.93	14.46			7.81***	62
Rater 2	63	46.34	14.83				
Film Text							
Rater 1	63	66.47	16.16			7.27***	62
Rater 2	63	50.20	14.64				
Prose Text							
Rater 1	53	54.72	19.36			2.90**	52
Rater 2	53	49.86	20.32				
Overall EFL Texts							
Rater 1	27			35.39	15.57	.38	26
Rater 2	27			34.77	15.30		
Opinions Text							
Rater 1	16			68.75	18.23	-.59	15
Rater 2	16			71.43	20.87		
Constitution Text							
Rater 1	22			46.75	17.73	.20	21
Rater 2	22			46.10	16.46		
Specialization Text							
Rater 1	7			60.71	28.35	1.92	6
Rater 2	7			46.43	9.45		

Table 4

Overall Summarizing Ability by EFL Course Level

	Hebrew Summaries				EFL Summaries			
	Intermediate		Advanced		Intermediate		Advanced	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Rater 1	8.89	2.14	8.04	2.14	7.59	2.50	4.30	2.00
ANOVA F Ratio	2.42						12.49**	
Rater 2	7.24	2.35	6.95	2.22	7.53***	2.70	4.10	.88
ANOVA F Ratio	1.55						15.03***	

***p<.001
**p<.01

Table 5

Summarizing Ability by EFL Course Level:
Crosstabulations of Differences at the Idea Level

Rater 1 - Idea 2 (Description)

		EFL Level	
	Pts.	Inter.	Adv.
Film Text in Hebrew	0	12	15
	1	25	11
	N =	37	26

d.f. 1
Chi-Square=3.98*

Rater 2 - Idea 2 (Description)

	EFL Level	
Pts.	Inter.	Adv.
0	12	16
1	25	10

d.f. 1
Chi-Square=5.24*

* p<.05

Rater 2 - Idea 7 (Exemplification)

	EFL Level	
Pts.	Inter.	Adv.
0	26	12
1	11	14
N =	37	26

d.f. 1
Chi-Square=3.71*

Rater 2 - Idea 3 (Justification for Constitution over other Laws)

	EFL Level	
Pts.	Inter.	Adv.
0	8	10
1	4	0
N =	12	10

Chi-Square=4.07*

* p<.05

Table 6

Correlations of Ratings for the Two Sets of Raters

EFL Texts

		Idea: 1 Purpose for Article	2 1st Sugges-tion	3 2nd Sug.	4 Ext of 2nd Sug.	5 3rd Sug.	6 4th Sug.	7	Overall
Opinions Text	Rater 1 with Rater 2	.19	.04	.45	-.09	.54*	.75***	--	.56*
	N=16								
	Idea:	1 Perspective	2	3 Justification	4 Details	5 Description	6 Descr.	7 Link	
Constitution Text	"	.59**	--	.47	.20	.91**	.46	-.05	.62**
	N=22								
	Idea:	1 Opener	2 General Ignorance	3 Ignor. in Field	4 Result of Mechanization				
Specialization Text	"	.17	1.00**	.47	.73				.94***
	N=7								

Correlation of ratings across all texts = .85*** (N=27)

Hebrew Texts

		Idea: 1	2	3	4	5	6	7	8	Overall
		De-scription	Descr.	Con-trast	Descr. Exempli-fication	Basic Descr.	De-tailed Descr.	Con-trast Exemplif.		
Film Text	Rater 1 w/ Rater 2	.55***	.84***	.23	.59***	.83***	.57***	.10	.20	.34*
	N=63									
		Contrast	De-scription	Descr.	Descr. Exemplification	Exemp.	Exemp.	Link		
Prose Text	"	.60**	.45**	.04	.06	.89***	.88***	.65***	.81***	
	N=53									

Correlation of ratings across all texts = .73*** (N=53)

* p<.05
** p<.01
*** p<.001

REFERENCES CITED

Bensoussan, M. & Kreindler, I. In press. Improving advanced reading comprehension in a foreign language: Summaries vs. short-answer questions. Journal of Research in Reading.

Brown, A.L. & Day, J.D. 1983. Macrorules for summarizing texts: The development of expertise. Journal of Verbal Learning and Verbal Behavior, 22, 1-14.

Chou Hare, V. & Borchardt, K.M. 1984. Direct instruction of summarization skills. Reading Research Quarterly, 20 (1), 62-78.

Cohen, A.D. 1984. On taking language tests: What the students report. Language Testing, 1 (1), 70-81.

Cohen, A.D. Forthcoming. The taking and rating of summarizing tasks on a foreign-language reading comprehension test. In Hill, C. & Parry, K. The test at the gate: Towards an ethnography of reading assessment. Cambridge: Cambridge University Press.

Cumming, A., Rebuffot, J., & Ledwell, M. 1989. Reading and summarizing challenging texts in first and second languages. Vancouver: Dept. of Language Education, U. of British Columbia.

Johns, A.M. 1985. Summary protocols of "underprepared" and "adept" university students: Replications and distortions of the original. Language Learning, 35 (4), 495-517.

Johns, A.M. & Mayes, P. Forthcoming. An analysis of summary protocols of university ESL students. Applied Linguistics.

Kintsch, W. & van Dijk, T.A. 1978. Toward a model of text comprehension and production. Psychological Review, 85 (5), 363-394.

Sarig, G. 1987. Testing meaning construction: Can we do it fairly? Language Testing, 6(1).

Sarig, G. 1988. Composing a study-summary: A reading-writing encounter. Ramat Aviv, Israel: Open University.

Shohamy, E. 1984. Does the testing method make a difference? The case of reading comprehension. Language Testing, 1 (2), 147-170.

Appendix A

Translation of the Elaborated Instructions:

Reading Text and Writing a Summary of the Text
Before you are four texts, two in Hebrew and two in English.

Instructions for Reading
Read each text so as to extract the most important points from it, that is, those points that contain the key sentences (for the given paragraph); or those points that the reader of the summary will be interested in reading.

Instructions for Writing the Summary (in Hebrew)
1. Reduce the information to main points only: avoid the inclusion of redundant information. Including this information will detract from the point!
2. Write briefly: the length of the summary is to be 80 words for the first passage and 120 words for the second.
3. Write the summary as a single passage: use connecting words to link the points together.
4. Do not translate literally: write the summary in your own words.
5. Write a draft first, and then copy it over legibly.

Translation of the Traditional Brief Instructions:

Reading the Text and Writing a Summary of It
Before you are four texts, two in Hebrew and two in English. You are to read each one so as to write a summary of it.

Instructions for Writing the Summary (in Hebrew)
1. Write briefly: the length of the summary is to be 80 words for the first passage and 120 words for the second.
2. Write a draft first, and then copy it over legibly.

Appendix B

Level of Agreement for Points Selected for Key

Hebrew Texts: "The Film Media -- from a Play Thing to an Art Form"
1. 5/9 2. 6/9 3. 7/9 4. 6/9 5. 5/9 6. 6/9 7. 8/9 8. 7/9

"The Problematicity of Modern Israeli Prose -- from Detachment to Continuity"
1. 8/9 2. 7/9 3. 7/9 4. 6/9 5. 9/9 6. 9/9 7. 8/9

"How to Avoid Foolish Opinions"
1. 8/9 2. 8/9 3. 7/9 4. 9/9 5. 9/9 6. 8/9 7. 8/9

"Modern Constitutions"
1. 8/9 2. 9/9 3. 6/9 4. 8/9 5. 7/9 6. 5/9 7. 8/9

"Specialization"
1. 7/9 2. 5/9 3. 9/9 4. 5/9